

Predicting Toxicity through a Computer Automated Structure Evaluation Program

by Gilles Klopman*

The computer automated structure evaluation program (CASE) has been extended to perform automatic quantitative structure-activity relationships (QSAR). Applications include the carcinogenicity of polycyclic aromatic hydrocarbons and of *N*-nitrosamines. Agreement with experiment is satisfactory.

The development and wide utilization of short-term assays (1) has led to the realization that a large number of chemicals present in the environment are potent bacterial mutagens and potential carcinogenic agents. Among these, polycyclic aromatic hydrocarbons (PAHs) (2), their nitro derivatives (3,4), a number of aromatic amines (5), and halogenated (6,7) as well as *N*-nitroso derivatives (8,9) are ubiquitous in the environment. Obviously, the presence of such biologically potent molecules in the air and in foods is cause for concern and it is not surprising to find that a major effort is underway to try to identify and eliminate these chemicals from the environment. However, while, as a class, these chemicals are suspicious, only a small fraction of them are really active and many show little or no activity. Thus, it is of importance to determine whether a more precise relationship between their structure and activity can be found.

The need for theoretical methods, capable of providing such relationships, is particularly important in view of the fact that it is inconceivable that every suspect chemical structure could be synthesized and tested in the laboratory.

Numerous methods, e.g., QSAR, pattern recognition, etc., aim at this goal, and most of them are discussed elsewhere in this volume. We will not review them here; instead, we will discuss our recently developed computer automated structure evaluation program (CASE) (10) which seems ideally suited for such a task and provides a pattern for the *a priori* recognition of potentially toxic molecules.

The Computer Automated Structure Evaluation Program (CASE)

One of the major difficulties in structure activity and quantitative structure-activity relationship (QSAR)

studies is the selection of appropriate molecular properties to be used as descriptors. Indeed, the success of a QSAR study often depends on the outcome of an intelligent and properly discriminant analysis of the factors that may be "determinant" in the observed activities.

The computer automated structure evaluation technique (10) addresses this problem and is used to select automatically the substructural units that are most appropriate to discriminate between active and inactive molecules. The method was described previously and consists of tabulating, for each molecule of a training set, the type of fragments that can be formed by breaking up the molecule in linear subunits consisting of 3 to 12 nonhydrogen atoms, together with the hydrogens attached to them. Each fragment generated by the breakup of a toxic molecule is labeled toxic, and each fragment generated by the breakup of a nontoxic molecule is labeled nontoxic. The fragments obtained from the full training set are collected and analyzed. Each type of fragment generated by the data base is evaluated on the premise that, if it is not related to toxic activity, it would be encountered randomly in toxic and nontoxic molecules. Any significant deviation from random distribution, at the 95% confidence level, is taken as an indication that the fragment is relevant to the observed toxicities. The method thus generates a finite set of substructural units presumed to be responsible for the observed toxicity of the molecules of the data set.

The choice of chemical substructural units as descriptors for activity makes a lot of sense to chemists who are used to relating chemical properties to functionalities and it is not surprising to find that much interest has centered around methodologies based on such descriptors (11-13). In most previous studies, though, only a finite number of preselected keys (11,12), consisting each of a well defined substructural unit, were considered. The problem with such an approach is that the selection of the keys was left to the imagination of the authors and to a large degree reflected their own bias.

*Chemistry Department, Case Western Reserve University, Cleveland, OH 44106

Furthermore, there was no guarantee that the selected keys were appropriate to handle the specific problem. In more recent developments, though, attempts were made at automating the selection of keys. The work of Chu (14) and Hodes (15,16) are examples of such developments and their programs are clearly related to our own methodology. The problem with these programs was that no causal relationship was established between the substructural descriptors and the activity of the molecules in the data base. Indeed, the keys were either too small or too restricted to be associated with the possible complex entities that give rise to biological functionality. Thus, the chemical analogy was lost because the activity, or lack of it, was related in a complex manner to the presence of a large number of "statistical" rather than "biological" keys.

This problem is addressed in our methodology. Indeed, we extended the number and size of potential keys but restricted them to biologically relevant ones by discriminant analysis. This resulted in the selection of only a few, but most appropriate, substructural descriptors (e.g., biophores) (17). To a degree, the program has intelligence, since it performs the painstaking task of selecting the appropriate descriptors usually performed by the researcher. Most importantly, though, it establishes causality and, in so doing, lays the ground to an understanding of the mechanism of actions of the toxic materials.

All functions of the program are performed automatically. The input consists merely of the *KLN* code (18) of the molecules of the training set and an evaluation on a scale of 1 to 9 of their activity. If, however, in response to the query for the activity of a new molecule, a question mark is entered, the program enters in its predictive mode and uses all the information that was fed to it previously to project the percentage chance that the new molecule is active. New data can be entered at any time and is immediately incorporated in the analysis of the problem. Thus, the program has learning capabilities as well.

We have already applied this methodology to a number of data bases and obtained good qualitative correlations for the carcinogenicity of polycyclic aromatic hydrocarbons and *N*-nitrosamines (10) as well as the genetic toxicity of nitroaromatic derivatives (17).

In our initial implementation of the method, biophores were identified on the grounds that they are capable of bestowing toxicity to a molecule; we now propose to carry the program one step further by evaluating the potency of each of the biophores. Such a technique had been pioneered by Free and Wilson (19) and is currently implemented in a number of computer packages. Our approach consists in tabulating the presence/absence of the appropriate keys in a training set of molecules and relating their presence via linear regression to the known quantitative activity of the molecules in which they are present.

Here, again, our implementation is totally automatic. Indeed, once the biophores have been identified, the

program proceeds to tabulate the presence/absence of these descriptors in the various molecules of the data base. It then proceeds to evaluate each of the descriptors in the context of a linear regression analysis and selects the most relevant one in a forward selection procedure.

In order to achieve the most accurate description, a new procedure for handling inactive molecules has been implemented. In this procedure, the inactive molecules are assigned a floating activity index whose value cannot exceed the value generally assigned to inactive molecules, but can vary downward as needed to produce the best linear plot. This permits us to include the inactive molecules in the regression analysis without the need of defining an "inactivity" ranking.

The result of this procedure is that the minimum number of descriptors, necessary to calculate the toxicity of the entire data set, is identified. The coefficients of the descriptors, which are seen as their potency, are automatically interfaced with the other data of the CASE program so that when a new compound not present in the initial data base is entered, in the predictive mode, a qualitative projection of the expected activity of the compound can also be made.

The program can clearly be used to study congeneric molecules. The interesting possibility exists that it can also address a diverse data base. Indeed, as long as the endpoint is well defined and the mechanism pretty well constant, nothing seems to prevent the possibility that enough different biophores are identified to accommodate the diversity of test compounds. Such a capability would present extraordinary opportunities to study a vast number of molecules. As yet, though, we have not built a sufficient data base to prove that this is indeed the case. Truly diverse data bases are rare; most of the time, they consist of combinations of data bases of congeneric molecules.

We describe below the application of the CASE program to two congeneric toxic data bases: the carcinogenicity of polycyclic aromatic hydrocarbons and that of *N*-nitrosamines. Then, in an initial effort to build a diverse data base, we have combined the two data bases and added a series of additional paraffinic molecules to see if a deterioration of the results takes place. To a degree, though, this is not a fair test, because the data bases are largely orthogonal. Our objective is to continue to incorporate additional molecules as we acquire appropriate data, and eventually build a truly diverse data base to be used to predict genetic toxicity of unknown molecules.

Results and Discussion

Polycyclic Aromatic Hydrocarbons

Polycyclic aromatic hydrocarbons (PAHs) are produced by combustion processes and are ubiquitous in the environment. Their carcinogenicity has been known for a long time and many attempts have been made at

Table 1. QSAR results for carcinogenic activity of polycyclic aromatic hydrocarbons.^{a,b}

Variable	Regression coefficient	Standard deviation
	(Constant = 1.847)	0.204
1	1.074	0.243
2	1.159	0.336
3	1.721	0.505
4	-0.354	0.169
5	-0.728	0.334
Standard deviation of residuals =		0.813
Index of determination (<i>R</i>) =		0.832

^aSubunit 1: C=C-CH=CH-CH=CH-C=CH-C=

Subunit 2: C=CH-CH=CH-CH=C-C=CH-CH=

Subunit 3: C=CH-CH=CH-C=CH-CH=C-C=CH

Subunit 4: CH=CH-CH=CH-CH=CH-C=CH-

Subunit 5: CH=CH-CH=CH-CH=CH-C=CH-CH=

CH-

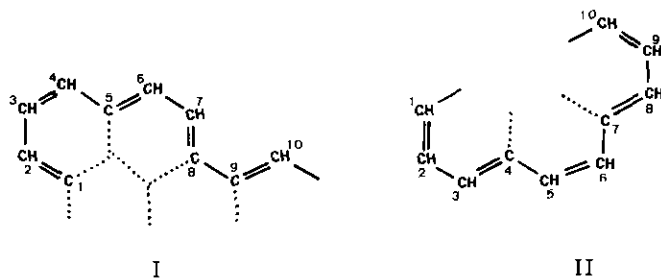
The dot (•) indicates that the previous atom bridges two rings.

^bNo. of molecules = 43; no. of false negatives = 7, of which 3 are marginal; no. of false positives = 0.

correlating their perceived genetic toxicity with their structure (5,20,21). It is now believed that their mechanism of action involves a "bay region" oxidation followed by alkylation of some genetic material (22).

We have previously (10) been able to obtain a good qualitative correlation between the carcinogenicity of these hydrocarbons, as reported by Dipple (23), and a few structural descriptors. We have now extended the treatment to a quantitative evaluation and have found the results shown in Table 1.

The data consisted of the 43 unsubstituted PAHs reported by Dipple as having no, marginal, moderate, or high carcinogenicity when painted on or injected in mice and rats. The data was reviewed and evaluated in a previous paper (10). As shown in Table 1, five descriptors were automatically selected by the program as being relevant to activity. Three of them are biophores, i.e., cause activity, while descriptors 4 and 5 are biophobes, i.e., prevent activity. While biophore 1, which describes essentially a "bay" region, is seen to be the most prevalent (i.e., it appears first in the list), biophore 3 appears to be the most potent (coefficient = 1.7). The latter represents the substructural unit (I).



The structure is clearly part of a benz(a)pyrene entity. Coupled with biophore 1, the bay region, it clearly bestows high activity to a molecule.

Table 2. QSAR results for carcinogenic activity of *N*-nitrosamines.

Variable	Regression coefficient	Standard deviation
	(Constant = 1.890)	0.3967
1	-2.427	0.5086
2	1.024	0.2017
3	1.748	0.4236
4		0.3479
Standard deviation of residuals =		1.166
Index of determination (<i>R</i>) =		0.734

^aSubunit 1: O=C-CH-

Subunit 2: O=N-N-CH₂-

Subunit 3: O=N-N-CH₃

Subunit 4: CH₂-CH₂-CH₂-CH₂-

^bNo. of molecules = 64; no. of false negative = 1; no. of false positive = 8.

Biophobe 4 defines an *L* region, clearly unfavorable, while biophobe 5 defines a kind of a *K* region (II), which is a rather surprising result. These effects, though, can be rationalized by considering that the two biophobes identify highly reactive regions that, if present, compete with the bay region for initial metabolism.

The correlation is satisfactory; *R* = 0.832 while the standard deviation is 0.813, i.e., less than one unit of activity. All inactive compounds were correctly assigned, but seven of the active compounds were not identified as carcinogens even though they were found experimentally to be moderately active.

N-Nitrosamines

N-Nitrosamines are believed to be produced, among other ways, by the action of nitrites, a food additive, on amines, a natural constituent of normal diet. Although no epidemiological data has been found that relates *N*-nitrosamines to carcinogenic events in humans, the molecules, in their pure form, are potent carcinogens in mice. The data base that we have used consists of data reported by Lijinski (24) on the incidence of esophagus and other cancers in rats. The animals were

Table 3. QSAR for carcinogenic activity of miscellaneous compounds.^{a,b}

Variable	Regression coefficient	Standard deviation
	(Constant = 1.8088)	0.1210
1	1.1105	0.0895
2	1.7800	0.2918
3	1.3330	0.2629
4	-2.5115	0.4110
5	1.1129	0.3797
Standard deviation of residuals =		0.969
Index of determination (<i>R</i>) =		0.800

^aSubunit 1: O=N-N-CH₂-

Subunit 2: O=N-N-CH₃

Subunit 3: C=CH-CH=CH-CH=CH-C=CH-C=

Subunit 4: O=C-CH-

Subunit 5: C=CH-CH=CH-CH=C-C=CH-CH=

A dot (•) indicates that the previous atom bridges two rings.

^bNo. of molecules = 127; no. of false negative = 9, of which 3 are marginal; no. of false positive = 8.

Table 4.

Activity				Activity			
No.	Molecule	Actual	Calculated	No.	Molecule	Actual	Calculated
1	Nitrosopiperidine	***	***	65	Propane		
2	2-Methylnitrosopiperidine	**	*	66	Butane		
3	3-Methylnitrosopiperidine	***	***	67	Isobutane		
4	4-Methylnitrosopiperidine	***	***	68	Cyclobutane		
5	2,6-Dimethylnitrosopiperidine			69	Pentane		
6	3,5-Dimethylnitrosopiperidine	*	***	70	Isopentane		
7	2,2,6,6-Tetramethylnitrosopiperidine			71	Neopentane		
8	4-Phenylnitrosopiperidine	**	***	72	Cyclopentane		
9	4-tert-Butylnitrosopiperidine		***	73	Hexane		
10	3-Hydroxynitrosopiperidine	***	***	74	2-Methylpentane		
11	4-Hydroxynitrosopiperidine	***	***	75	3-Methylpentane		
12	4-Ketonitrosopiperidine	***	***	76	2,2-Dimethylbutane		
13	2-Carboxynitrosopiperidine			77	Cyclohexane		
14	4-Carboxynitrosopiperidine			78	Heptane		
15	4-Chloronitrosopiperidine	****	***	79	2-Methylhexane		
16	3,4-Dichloronitrosopiperidine	****	***	80	3-Methylhexane		
17	3,4-Dibromonitrosopiperidine	****	***	81	2,2-Dimethylpentane		
18	Nitroso-1,2,3,6-tetrahydropyridine	****	***	82	2,2,3-Trimethylbutane		
19	Nitrosoguvacoline			83	Cycloheptane		
20	Nitrosomethylphenidate			84	Octane		
21	Nitrosopyrrolidine	**	***	85	Benzene		
22	2,5-Dimethylnitrosopyrrolidine			86	Naphthalene		
23	3,4-Dichloronitrosopyrrolidine	****	***	87	Anthracene		
24	2-Carboxynitrosopyrrolidine			88	Phenanthrene		
25	2-Carboxy-4-hydroxynitrosopyrrolidine			89	Pyrene		
26	Nitroso-3-pyrroline	**	***	90	Naphthacene		
27	Nitrosomorpholine	***	***	91	Triphenylene		
28	2,6-Dimethylmorpholine	****	***	92	Chrysene	*	
29	Nitrosothiomorpholine	**	***	93	Benz(a)anthracene	*	
30	Nitrosophenmetrazine		*	94	Benz(c)phenanthrene	**	*
31	Dinitrosopiperazine	***	****	95	Benz(a)pyrene	***	**
32	2-Methyldinitrosopiperazine	****	****	96	Benz(e)pyrene		
33	2,5-Dimethyldinitrosopiperazine	***	***	97	Perylene		
34	2,6-Dimethyldinitrosopiperazine	****	***	98	Pentacene		
35	2,3,5,6-Tetramethyldinitrosopiperazine			99	Benz(a)naphthacene		
36	Dinitrosohomopiperazine	****	****	100	Benz(o)chrysene		
37	Nitrosopiperazine		***	101	Benz(c)chrysene	**	*
38	4-Methylnitrosopiperazine		***	102	Benz(g)chrysene	**	*
39	Nitrosoazetidine	***	***	103	Picene		
40	Nitrosohexamethyleneimine	****	***	104	Dibenz(ac)anthracene	*	
41	Nitrosoheptamethyleneimine	****	***	105	Dibenz(ah)anthracene	**	
42	Nitrosooctamethyleneimine	***	***	106	Dibenz(aj)anthracene	**	
43	Nitrosododecamethyleneimine	**	***	107	Dibenz(cg)phenanthrene		
44	Dimethylnitrosamine	****	****	108	Dibenz(bg)phenanthrene		
45	Diethylnitrosamine	****	***	109	Pentaphene		
46	Bis(2-chloro)diethylnitrosamine	***	***	110	Anthranthrene		
47	Bis(2-cyano)diethylnitrosamine		***	111	Benz(ghi)perylene	**	
48	Bis(2-methoxy)diethylnitrosamine	***	***	112	Dibenz(ae)pyrene	***	**
49	Bis(2-ethoxy)diethylnitrosamine	***	***	113	Dibenz(al)pyrene	***	****
50	Bis(2,2-diethoxy)diethylnitrosamine		***	114	Dibenz(ah)pyrene	***	***
51	Diisopropylnitrosamine	**		115	Dibenz(ai)pyrene	***	***
52	Di-sec-butylnitrosamine			116	Dibenz(el)pyrene		
53	Di-n-propylnitrosamine	****	***	117	Naphto(23a)pyrene	**	
54	Bis(2-hydroxy)-n-propylnitrosamine	**	***	118	Naphth(23e)pyrene		
55	Bis(2-oxo)-n-propylnitrosamine	***	***	119	Dibenz(bk)chrysene		
56	Nitrosomethoxymethylamine		**	120	Dibenz(aj)naphthacene		
57	Nitrosomethylethylamine	**	***	121	Dibenz(ao)naphthacene	**	
58	Nitrosomethylundecylamine	***	***	122	Anth(12a)anthracene		
59	Nitrosomethyldodecylamine	***	***	123	Benz(c)pentaphene		
60	Nitrosodi-n-octylamine		***	124	Benz(b)pentaphene		
61	Nitrosomethyl-2-phenylethylamine	****	***	125	Napht(12a)triphenylene		
62	Nitrosomethylneopentylamine	****	***	126	Hexacene		
63	Nitrosomethylphenylamine	***	**	127	Tribenz(aei)pyrene	**	***
64	Nitrosomethylcyclohexylamine	***	**				

fed a daily dose of various *N*-nitrosamines as 5% of their total diet. The end point is a rating from 1 to 4 indicating the incidence of cancer in these rats.

Here also, we had previous experience with the data (25), having used part of it to evaluate qualitatively the toxicity of the cyclic compounds. We have now extended the data base to include 69 *N*-nitrosamines, both cyclic and acyclic. The results appear in Table 2. The program selected the four descriptors shown below Table 2.

All except one are biophores, reflecting the large proportion of *N*-nitrosamines that showed activity. The only biophobe, CH-C=O , indicates that the presence of a carbonyl group in any of the molecules prevented activity. Biophores 2 and 3 indicate that, to be carcinogenic, the carbon atoms alpha to the *N*-nitroso group must bear at least two hydrogen atoms, a fact that is compatible with the preferred mechanism of action of these molecules (26). Biophore 3 is the most potent (coefficient = 1.7), indicating that methylation is probably the most toxic of the possible alkylation events.

The program recognized all but one active molecule; however, it predicted eight of the inactive ones to be active. This may be indicative of the fact that some additional biophobe might not have been identified properly. An alternative explanation, though, could also be that some additional important descriptor, for example the partition coefficient, plays an important role in the activity of the molecules.

Miscellaneous Data Base

In this data base, we included both the data for the PAHs and the *N*-nitrosamines. We also added 20 miscellaneous paraffins, believed to be inert, and submitted the total 127 compounds to CASE analysis. To a degree, the three data bases that constitute this miscellaneous group are not consistent, since the endpoint had been evaluated differently in each of them. Nevertheless, the property they represent is sufficiently well defined to warrant an attempt at normalization. The program picked up five descriptors of which four were biophores and one a biophobe. The results appear in Table 3.

As can be seen, all descriptors were already represented in the individual QSARs. Thus the data base (Table 4) is not truly diverse, since the structures of the three categories of compounds are largely orthogonal. Nevertheless, as a first step towards the development of a QSAR for diverse compounds, the results are of some significance.

The index of correlation R was found to be 0.800 while the standard deviation of residuals is now close to one activity unit, i.e., 0.969. The number of incorrect predictions is 17, i.e., about 15% of the total number of molecules in the data base. This is probably the lower limit of what can be expected from predictions derived from such calculations. Close examination of the data reveals that the errors in the global QSAR strongly parallel those found in the individual ones. This again

is an indication of the orthogonality of the congeneric data bases that were used to make it up.

Conclusion

The method works well and appears useful to correlate and possibly predict the activity of congeneric molecules. Further work is needed to assess the potential of this method for general application to diverse data sets.

The process is totally automatic and the QSAR are obtained merely by requesting such evaluation to be made. The nature of the descriptors is clearly related to the activity of the molecule, thus providing the impetus to initiate other types of QSAR applications, based for example, on the quantum mechanical indices of the atoms of the substructures that are selected as being relevant to activity. We plan to investigate these avenues in the future.

The author wishes to thank Professor H. Rosenkranz for stimulating this work, for helping with numerous suggestions, and providing much of the needed biological data. The author is also grateful to the Environmental Protection Agency for their financial support for the study of structure-activity relationships in potential carcinogens (Grant RS100623).

The data base for nitrosamines was provided by Dr. W. Lijinsky, Chemical Carcinogenesis Program, Frederick Cancer Research Center, Frederick, MD 21701 (USA) (1978).

REFERENCES

1. Ames, B. N., McCann, J., and Yamasaki, E. Methods for detecting carcinogens and mutagens with the *Salmonella*/mammalian-microsome mutagenicity test. *Mutat. Res.* 31: 347-363 (1975).
2. Schuetzle, D., Riley, T. L., Prater, T. J., Salmeen, I., and Harvey, T. M. The identification of mutagenic chemical species in air particulate samples. *Environ. Sci.* 7: 259-280 (1982).
3. Jager, J. Detection and characterization of nitro derivatives of some polycyclic aromatic hydrocarbons by fluorescence quenching after thin-layer chromatography: application to air pollution analysis. *J. Chromatogr.* 152: 575-578 (1978).
4. Pitts, J. N. Jr., Van Cauwenberghe, K. A., Grosjean, D., Schmid, J. P., Fitz, D. R., Belser, W. L. Jr., Knudsen, G. B., and Hunds, P. M. Atmospheric reactions of polycyclic aromatic hydrocarbons: facile formation of mutagenic nitro derivatives. *Science* 202: 519 (1979).
5. Arcos, J. C., and Argus, M. F. Molecular geometry and carcinogenic activity of aromatic compounds. New perspectives. *Advan. Cancer Res.* 11: 305-471 (1968).
6. McConnell, E. E. Halogenated biphenyls, terphenyls, naphthalenes, dibenzodioxins and related products. Acute and chronic toxicity, carcinogenesis, reproduction, teratogenesis and mutagenesis in animals. In: *Topics Environ. Health*, 1980: 109-150.
7. Ware, S. A., and West, W. L. Investigation of selected potential environmental contaminants: halogenated benzenes. U.S. NTIS, PB Rep., PB-273206 (1977).
8. Hecht, S. S., McCoy, G. D., Chen, C. B., and Hoffmann, D. On the metabolism of cyclic nitrosamines. In: *N-Nitroso Compounds* (R. A. Scanlan and S. R. Tannenbaum, Eds.), ACS Symp. Ser. American Chemical Society, Washington, DC, 1981, pp. 49-75.
9. Wishnok, J. S. On effects of structure on the carcinogenic behavior of nitrosamines. In *N-Nitroso Compounds* (R. A. Scanlan

- and S. R. Tannenbaum, Eds.), ACS Symp. Ser. American Chemical Society, Washington, DC, 1981, pp. 77-87.
10. Klopman, G. Artificial intelligence approach to structure-activity studies. *J. Am. Chem. Soc.* 106: 7315-7321 (1984).
 11. Enslein, K., and Craig, P. N. Carcinogenesis; a predictive structure-activity model. *J. Toxicol. Environ. Health* 10: 521-530 (1982).
 12. Jurs, P. C., Chou, J. T., and Yuan, M. Studies of chemical structure-biological activity relations using pattern recognition. In: *Computer Assisted Drug Design* (E. C. Olson and R. E. Christoffersen, Eds.), ACS Symp. Ser. No. 112, American Chemical Society, Washington, DC, 1979, pp. 103-129.
 13. Hodes, L. Computer-aided selection of novel antitumor drugs for animal screening. In: *Computer Assisted Drug Design* (E. C. Olson and R. E. Christoffersen, Eds.), ACS Symp. Ser. 112, American Chemical Society, Washington, DC, 1979, pp. 583-602.
 14. Chu, K. C., Feldmann, R. J., Shapiro, M. B., Hazard, G. F., Jr., and Geran, R. I. Pattern recognition and structure-activity relation studies. Computer-assisted prediction of antitumor activity in structurally diverse drugs in an experimental mouse brain tumor system. *J. Med. Chem.* 18: 539-545 (1975).
 15. Hodes, L., Hazard, G. F., Geran, R. I., and Richman, S. A statistical-heuristic method for automated selection of drugs for screening. *J. Med. Chem.* 20: 469-475 (1977).
 16. Tinker, J. F. A computerized structure-activity correlation program for relating bacterial mutagenesis activity to chemical structure. *J. Comput. Chem.* 2: 231-243 (1981).
 17. Klopman, G., and Rosenkranz, H. Structural requirements for the mutagenicity of environmental nitroarenes. *Mutat. Res.* 126: 227-238 (1984).
 18. Klopman, G., and McGonigal, M. Computer simulation of physical-chemical properties of organic molecules. 1. Molecular system identification. *J. Chem. Inf. Comput. Sci.* 21: 48-52 (1981).
 19. Free, S. M., Jr., and Wilson, J. W. A mathematical contribution to structure-activity studies. *J. Med. Chem.* 7: 395-399 (1964).
 20. Pullman, A., and Pullman, B. Electronic structure and carcinogenic activity of aromatic molecules. *Adv. Cancer Res.* 3: 117-169 (1955).
 21. Jerina, D. M., and Lehr, R. E. The bay-region theory: a quantum mechanical approach to aromatic hydrocarbon-induced carcinogenicity. In: *Microsomes and Drug Oxidation* (V. Ulrich, I. Roots, A. G. Hildebrandt, R. W. Eatabrook and A. H. Conney, Eds.), Pergamon Press, Oxford, 1977, pp. 709-720.
 22. Borgen, A., Darvey, H., Castagnoli, N., Crocker, T. T., Rasmussen, R. E., and Wang, I. Y. Metabolic conversion of benzo[a]pyrene by Syrian hamster liver microsomes and binding of metabolites to deoxyribonucleic acid. *J. Med. Chem.* 16: 502-506 (1973).
 23. Dipple, A. Polynuclear aromatic carcinogens. In: *Chemical Carcinogens* (C. E. Searle, Ed.), ACS Monograph, American Chemical Society, Washington, DC, 1976, pp. 173-245.
 24. Lijinsky, W., and Taylor, H. W. The effect of substituents on the carcinogenicity of *N*-nitrosopyrrolidine in Sprague-Dawley rats. *Cancer Res.* 36: 1988-1990 (1976).
 25. Petit, B., Potenzzone, R., Hopfinger, A. J., Klopman, G., and Shapiro, M. A hierarchical QSAR molecular structure calculator applied to a carcinogenic nitrosamine data base. In: *Computer Assisted Drug Design* (E. C. Olson and R. E. Christoffersen, Eds.), ACS Symp. Ser. American Chemical Society, Washington, DC, 1979, pp. 553-581.
 26. Andreozzi, P., Klopman, G., and Hopfinger, A. J. Theoretical study of *N*-nitrosamines and their presumed proximate carcinogens. *Can. Biochem. Biophys.* 4: 209-219 (1980).